ELEC 573: Network Science and Analytics

Fall 2019

Instructor: Santiago Segarra (segarra@rice.edu), office: Duncan Hall 2047, office hours: TBD.

Time and location: Thursday 04:00-06:50 PM, Location TBD

Short description: This course provides an introduction to complex networks, their structure, and function, with examples from engineering, biology, and social sciences. Topics include spectral graph theory, notions of centrality, community detection, random graph models, inference in networks, opinion dynamics, and contagion phenomena. Our main goal is to study network structures and how they can be leveraged to better understand data defined on them.

Prerequisites: Linear algebra, probability and statistics, and basic ability to program in Python.

Learning outcomes: This course will provide students with the mathematical tools and computational training to understand large-scale networks in the current era of Big Data. It is our objective that after taking this course the students will be able to model a variety of systems as networks and will be able to compute and interpret structural descriptors of these models. Moreover, students will be able to make statistical predictions of unobserved portions of the network systems and model the potential evolution of processes on the network. Finally, our objective is that students will be able to use networks as a universal language to communicate across different application domains, providing the opportunity to implement engineering solutions to seemingly unrelated fields.

Grading policy: Homework (40%) and final project (60%).

Homework: The course will have three homework sets, each of them having a two-week period between its publication and its due date. Every homework will be graded out of 100 points. At the end of the course, your homework set with the lowest score will weigh 10% towards the final course grade while the other two will weigh 15% each (thus adding up to the total of 40%).

Final project: You will work in groups on a research project where you can investigate and apply state-of-the-art network analysis tools and algorithms to an application of your preference. Keep in mind that your project is an in-depth study of a specific topic; it is not just a summary of a few research papers. We will give more details about what constitutes an admissible project throughout the course, as well as a list of topics of potential interest.

The project will have three associated due dates with specific deliverables and grade weights:
i) Project proposal (10%): Brief summary of what you plan to do for your project.
ii) Progress report (15%): First (incomplete) draft of your final report, though naturally shorter and most likely without your major results. Should serve as a skeleton for the final report and as a checkpoint to assess feasibility.
iii) Final report and presentation (35%): Should provide a clear and detailed description of what you did, what results you obtained, and what you have learned and concluded from your work. Apart from the written report, you will prepare a short presentation for the rest of the class.

Collaboration policy and the Rice Honor Code: We encourage that you work together whenever possible: problem sets and general discussion of the material. Keep in mind, however, that for the problem sets the solutions you hand in should reflect your *own* understanding of the class material, and *should be written solely by you*. It is *not* acceptable to copy a solution (theory or code) that somebody else has written.

In this course, all students will be held to the standards of the Rice Honor Code, a code that you pledged to honor when you matriculated at this institution. If you are unfamiliar with the details of this code and how it is administered, you should consult the Honor System Handbook at http://honor.rice.edu/honor-system-handbook/. This handbook outlines the University's expectations for the integrity of your academic work, the procedures for resolving alleged violations of those expectations, and the rights and responsibilities of students and faculty members throughout the process.

Attendance policy: Student are encouraged (but not required) to attend lectures.

Course syllabus: Here we present a high-level description of the content units that will be covered in the course:

i) Introduction to the course (0.5 lecture): We first cover course specifics and logistics: office hours, homework, final project, bibliography, etc. We then introduce the notion of network and present a general notion of Network Science as a cross-disciplinary field. We will motivate this via several examples and highlight the associated challenges.

ii) Introduction to graph theory (1 lectures): We cover basic notions and definitions related to undirected and directed graphs: vertices, edges, simple graphs, weighted graphs, neighborhoods, degree, path, cycle. Also, we introduce concepts from algebraic and spectral graph theory, such as adjacency, incidence, and Laplacian matrices.

iii) Centrality measures (1 lectures): We introduce the notion of node centrality as a way to measure the importance of a node within the network. We compare classical measures such as degree, closeness, betweenness, eigenvector, and Katz centrality. We then continue with link analysis algorithms for web search and describe the PageRank algorithm for ranking nodes in graphs in the context of random walks.

iv) Community detection (1.5 lectures): We first introduce the notions of cohesive subgroups, clustering coefficient, homophily, and modularity. We then discuss the famous 'strength of weak ties' hypothesis from sociology and its relation with communities in graphs. We then

formulate the problem of community detection as a graph partitioning problem and study several algorithms in detail including spectral and hierarchical clustering.

v) Network models (2 lectures): We discuss the notions of degree distributions, power laws, and popularity in networks. These naturally lead to the definition of important classes of network graph models: Erdős-Rényi, small-world, and preferential attachment. The emphasis is on model construction, simulation, and inference of model parameters.

vi) Inference of network topologies and features (2.5 lectures): We first introduce the fundamental problem of estimating a network summary characteristic from sampled graph data, and present the Horvitz-Thompson estimator as well as the capture-recapture method. Several network graph sampling designs are then discussed. We then analyze in quite detail the problem of network topology inference, where a portion of the network is unobserved and we wish to infer it from measurements.

vii) Inference of network processes (1 lectures): We present the problem of prediction static processes defined on networks. We start with the simple non-parametric method of nearest-neighbor prediction, and then transition to model-based approaches that describe vertex attributes in terms of the given network graph structure. Specifically, we study Markov random field models and graph-kernel regressions.

viii) Epidemics and contagion (1 lectures): We study the spread of epidemics in networks. Starting from branching processes and classical epidemiology we then study the network-based SI, SIR, and SIS models of diffusion. We analyze transient properties of these processes as well as the problem of inferring the source of an epidemic spread.

xi) Graph signal processing and deep learning architectures for network data (2 lectures): We provide a first introduction to geometric deep learning and study how the basic deep learning architectures (convolutional neural networks, autoencoders) generalize to graph-based data. We will analyze data-driven methods to solve graph problems such as link-prediction, semi-supervised learning, and graph generation.

Recommended texts: There is no *required* text. The main material is contained in the course slides and supported by in-class derivations on the board. However, as additional references and for the curious students that want to expand on these topics, we recommend the following books:

[1] Newman, M. (2010). *Networks: an Introduction*. Oxford university press.

[2] Kolaczyk, E. (2009). *Statistical Analysis of Network Data: Methods and Models*. Springer Series in Statistics.

[3] Easley, D. and Kleinberg, J. (2010). *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press.

Disability support services: If you have a documented disability or other condition that may affect academic performance you should: 1) make sure this documentation is on file with Disability Support Services (Allen Center, Room 111 / adarice@rice.edu / x5841) to determine the accommodations you need; and 2) talk with me to discuss your accommodation needs.

Syllabus change policy: This syllabus is a general guide for the course and is subject to change with advanced notice.